Orthogonal Subspace Learning for Language Model Continual Learning

Abstract

Benefiting from massive corpora and advanced hardware, large language models (LLMs) exhibit remarkable capabilities in language understanding and generation. However, their performance degrades in scenarios where multiple tasks are encountered sequentially, also known as catastrophic forgetting. In this paper, we propose orthogonal low-rank adaptation (O-LoRA), a simple and efficient approach for continual learning in language models, effectively mitigating catastrophic forgetting while learning new tasks. Specifically, O-LoRA learns tasks in different (low-rank) vector subspaces that are kept orthogonal to each other in order to minimize interference. Our method induces only marginal additional parameter costs and requires no user data storage for replay. Experimental results on continual learning benchmarks show that our method outperforms state-of-the-art methods. Furthermore, compared to previous approaches, our method excels in preserving the generalization ability of LLMs on unseen tasks.